

NDIIPP/NDSA meeting, Washington, July 19-21, 2011

# JHOVE2 from Project to Community

— *or, is there life after NDIIPP?*

Stephen Abrams  
*California Digital Library*

Sheila Morrissey  
*Portico*

*JHOVE2 Project Team*

*“Tell me about yourself...”*



## Characterization

- The automated determination of the intrinsic and extrinsic properties of a formatted object
  - Identification *Determining the **presumptive** format of a digital object based on suggestive extrinsic hints and intrinsic signatures*
  - Feature extraction *Reporting the **intrinsic** properties of an object significant for classification, analysis, and planning*
  - Validation *Determining the level of **conformance** to the normative requirements of a format's authoritative specification*
  - Assessment *Determining the level of **acceptability** for a specific purpose on the basis of locally-defined policy rules*



“We report, you decide”

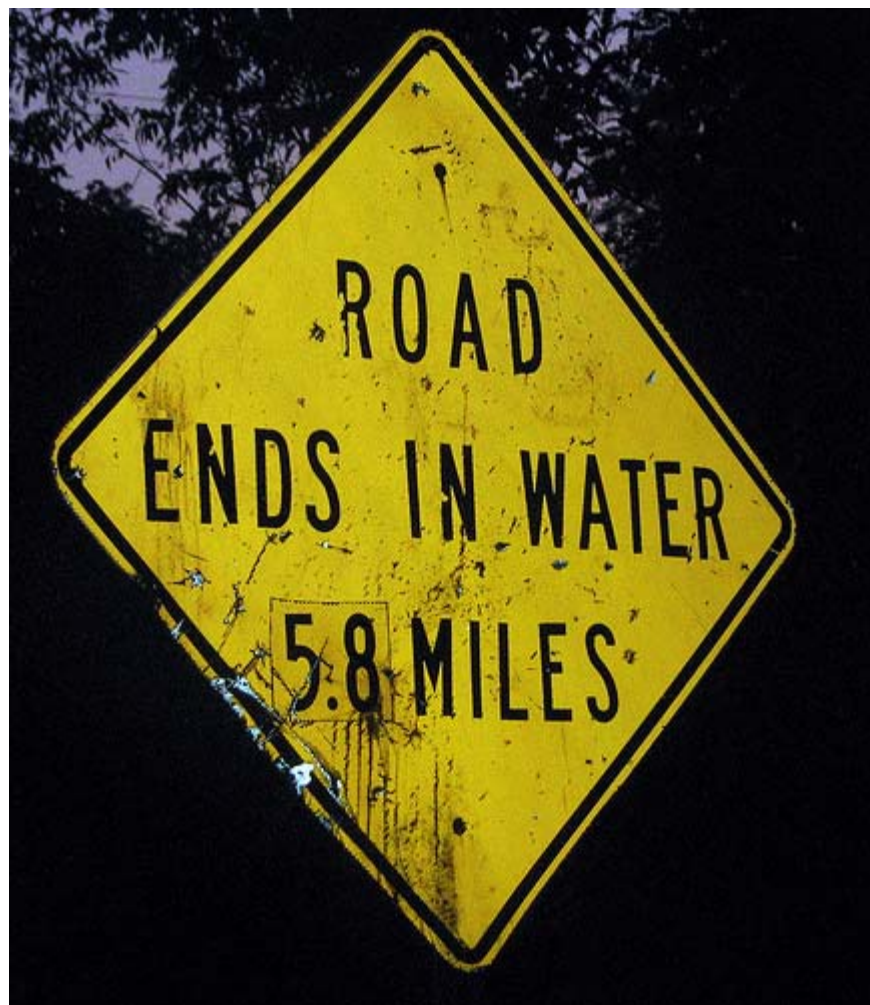
## *JHOVE2 project*

- A two year project to develop a next-generation open source framework and application for format-aware characterization
- Collaboration between the California Digital Library (CDL), Portico, and Stanford University
- Funded by the Library of Congress as part of its National Digital Information Infrastructure Preservation Program (NDIIPP)

## *Project goals*

- Make the code easy to understand and extend
  - Design for reliance on community enthusiasm and commitment
    - It's the modules, stupid
    - It's not just us, it's everyone
  - It's already working!
    - Wegener Institute development of modules for NetCDL and Grib
    - BnF / Atos development of modules for Gzip and ARC
    - BnF / NetArchive / ONB development of module for WARC
    - Scape project

— *but what happens next?*

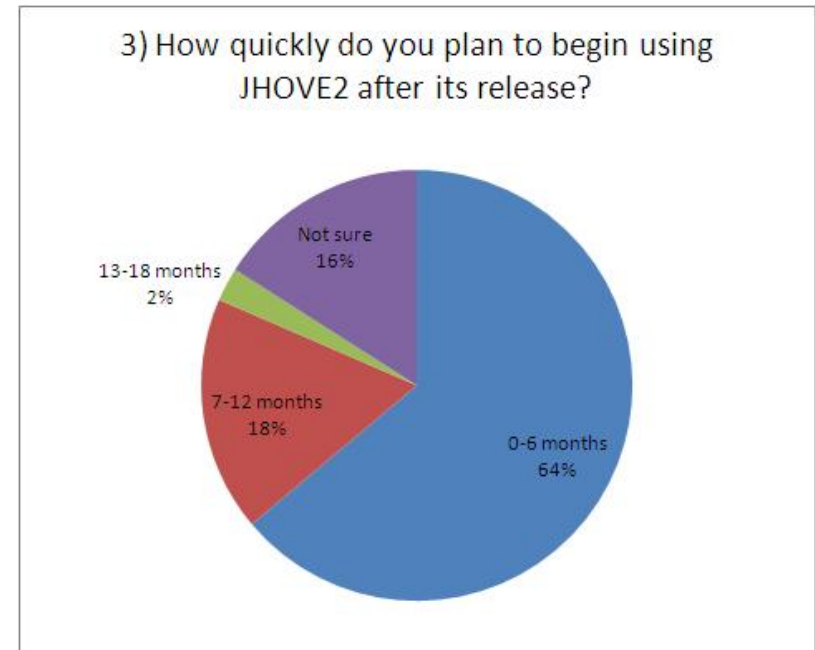
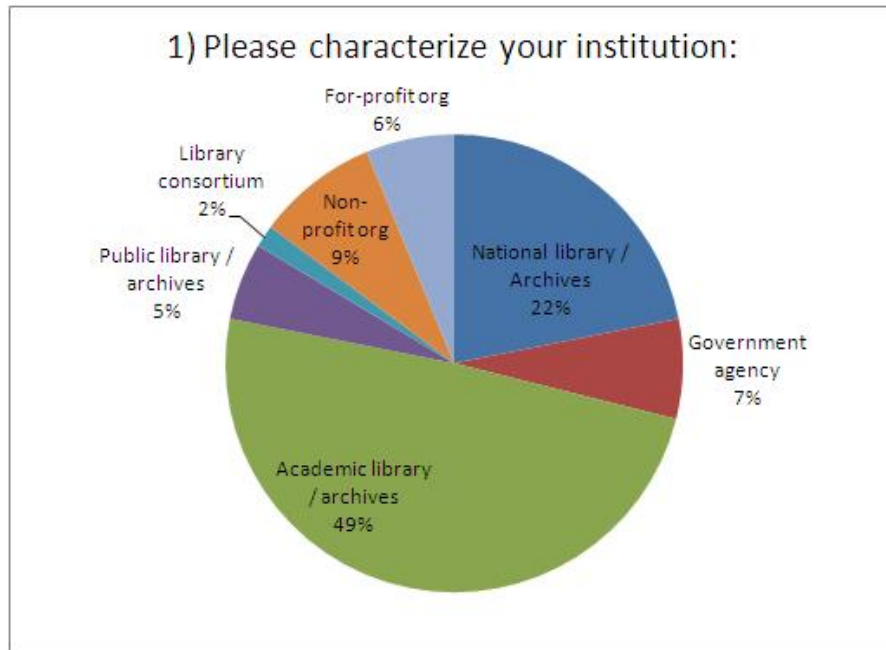


## *Post-project planning*

- Production 2.0.0 release in April 2011
  - Over 200 downloads to date
- Additional releases scheduled for later in 2011
  - 2.1.0 ARC and Gzip integration, JPEG 2000
  - 2.2.0 PDF, Zip (*fully validating*)
- Project partners are committed to provide ongoing, self-funded support for three years
- By year 4, project partners expect to have transitioned long-term responsibility to ...?

## User-driven priorities

- Planned activities are based on user survey results
  - 145 respondents, 88 institutions, 23 countries

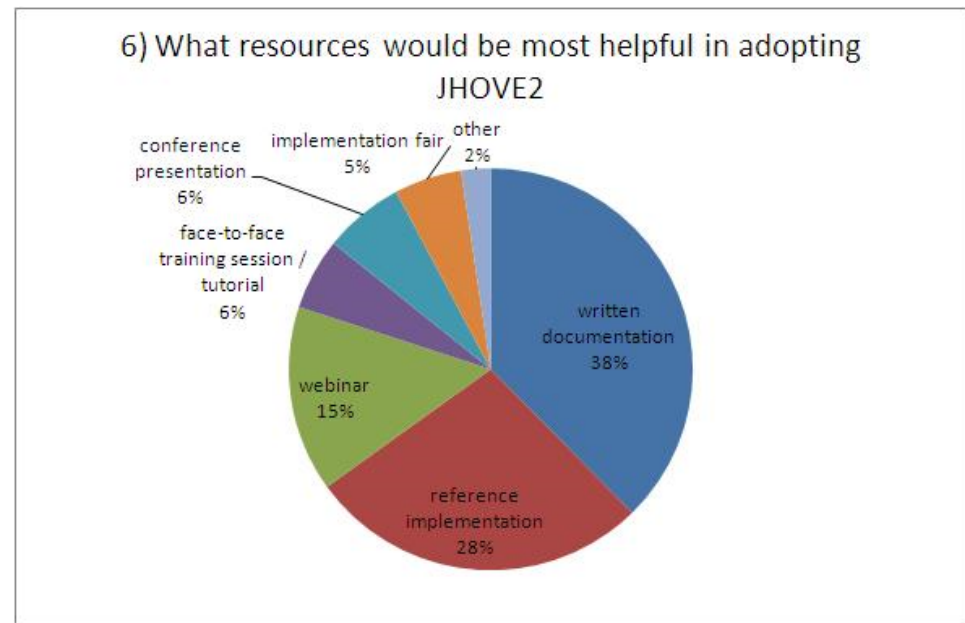
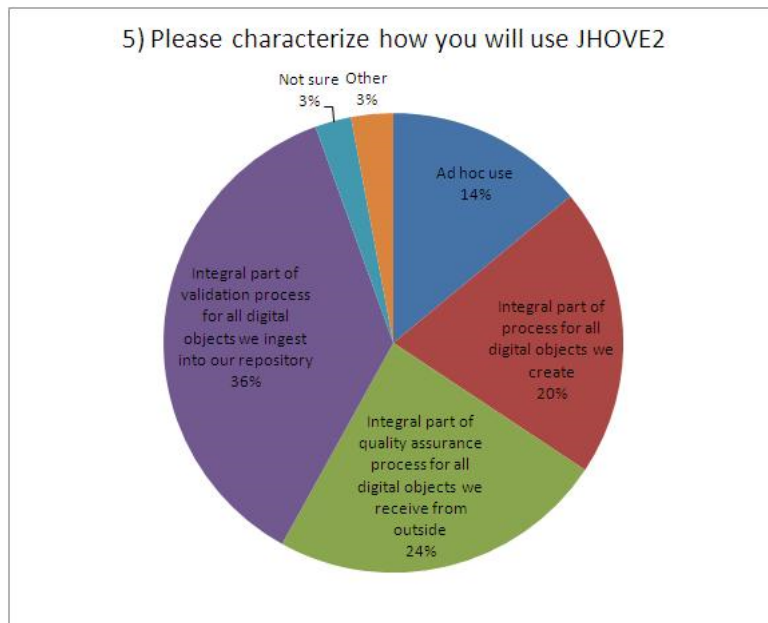


Full results available at [https://bitbucket.org/jhove2/main/wiki/User\\_survey](https://bitbucket.org/jhove2/main/wiki/User_survey)



## User-driven priorities

- Planned activities are based on user survey results
  - 145 respondents, 88 institutions, 23 countries



Full results available at [https://bitbucket.org/jhove2/main/wiki/User\\_survey](https://bitbucket.org/jhove2/main/wiki/User_survey)

## *Sustainable activities*

- Support and maintain the core JHOVE2 code
  - Bitbucket / mercurial / maven
- Provide training on integration and use
  - Complete documentation package, workshops
- Solicit and support 3<sup>rd</sup> party module development
- Solicit and support integration with other systems
- Establish a lightweight community structure to guide and foster JHOVE2 technical development
- Suggestions welcome, volunteers encouraged

# *Community*

- Steering group
  - Three project partners, Library of Congress, other contributors as appropriate, all dedicated to sustaining the project and codebase
- Advisory group
  - Providing strategic input and resources for maintenance and enhancement
- Committers group
  - Core developers, advancing the core codebase, integrating contributions, and managing releases
- Open source community
  - JHOVE2 users and code contributors

*Help!*



*For more information...*

<http://jhove2.org/>

[JHOVE2-Announce-L@listserv.ucop.edu](mailto:JHOVE2-Announce-L@listserv.ucop.edu)

[JHOVE2-Techtalk-L@listserv.ucop.edu](mailto:JHOVE2-Techtalk-L@listserv.ucop.edu)

#### CDL/UC3

*Stephen Abrams*

*Patricia Cruse*

*John Kunze*

*Isaac Rabinovitch*

*Marisa Strong*

*Perry Willett*

#### Stanford University

*Richard Anderson*

*Tom Cramer*

*Hannah Frost*

#### Portico

*John Meyer*

*Sheila Morrissey*

#### Library of Congress

*Martha Anderson*

*Justin Littman*

#### With help from

*Walter Henry*

*Nancy Hoebelheinrich*

*Keith Johnson*

*Evan Owens*

#### Advisory Board

*Deutsche Nationalbibliothek*

*DSpace / MIT*

*Ex Libris*

*Fedora Commons / Rutgers*

*Florida Center for Library Automation*

*Harvard University*

*Koninklijke Bibliotheek*

*National Archives [UK]*

*National Archives [US]*

*National Library of Australia*

*National Library of New Zealand*

*Planets / Universität zu Köln*

*Tessella*